applied sciences

MDPI

Article

Diffuse-Field Equalisation of Binaural Ambisonic Rendering

Thomas McKenzie *, Damian T. Murphy and Gavin Kearney

Audio Lab, Communication Technologies Research Group, Department of Electronic Engineering, University of York, York YO10 5DD, UK; damian.murphy@york.ac.uk (D.T.M.); gavin.kearney@york.ac.uk (G.K.)

* Correspondence: ttm507@york.ac.uk; Tel.: +44-1904-32-4233

Received: 9 October 2018; Accepted: 13 October 2018; Published: 17 October 2018

Abstract: Ambisonics has enjoyed a recent resurgence in popularity due to virtual reality applications. Low order Ambisonic reproduction is inherently inaccurate at high frequencies, which causes poor timbre and height localisation. Diffuse-Field Equalisation (DFE), the theory of removing direction-independent frequency response, is applied to binaural (over headphones) Ambisonic rendering to address high-frequency reproduction. DFE of Ambisonics is evaluated by comparing binaural Ambisonic rendering to direct convolution via head-related impulse responses (HRIRs) in three ways: spectral difference, predicted sagittal plane localisation and perceptual listening tests on timbre. Results show DFE successfully improves frequency reproduction of binaural Ambisonic rendering for the majority of sound source locations, as well as the limitations of the technique, and set the basis for further research in the field.

Keywords: ambisonics; binaural; equalisation

1. Introduction

Ambisonics is a three-dimensional spatial audio approach based on the spatial sampling and reconstruction of a sound field using spherical harmonics, first introduced by Michael Gerzon in the 1970s [1,2]. Ambisonic sound fields can be recorded using multiple capsule microphone arrays or encoded using point sources, and reproduced over multiple loudspeaker arrays or over headphones. In Ambisonic playback, assuming free-field reproduction conditions and plane wave propagation, the reproduced sound field is accurate only for the region of a head in the centre of the loudspeaker array, for a frequency range up to the spatial aliasing frequency f_{alias}. This is due to the inherent limited spatial accuracy of recording and reproduction of a physical sound field with a finite number of transducers.

Above f_{alias}, timbral inconsistencies exist due to comb filtering from the summation of coherent loudspeaker signals with multiple delay paths to the ears, which are not situated in the exact centre of the loudspeaker array. In practice this reduces the accuracy of reproduced timbre and height cues. By increasing the order of Ambisonics, which requires more microphones and encoded channels in production and storage and more loudspeakers in reproduction, f_{alias} rises improving both localisation and timbre, though for all practical Ambisonic rendering systems at present f_{alias} is still very much within the human hearing range.

Horizontal localisation accuracy can be improved above f_{alias} by employing different decoding methods for high frequencies [3], though this does not improve timbre. As accurate timbral reproduction has been shown to be more important than localisation when measuring the authenticity of a spatial audio experience [4,5], it is essential to address timbre above f_{alias} in Ambisonic reproduction.

The traditional method for rendering Ambisonic sound fields over headphones is to decode the sound field to a specified loudspeaker configuration (as in loudspeaker rendering), and then take

the sum of a convolution of the loudspeaker signals with head-related impulse responses (HRIRs) corresponding to each loudspeaker's position [6,7] (often referred to as the virtual loudspeaker approach). Recent methods for binaural rendering of Ambisonics have moved away from the virtual loudspeaker approach and instead focused on order truncation of a spatially continuous spherical harmonic (SH) represented head-related transfer functions (HRTF) data set [8–10], which through pre-processing techniques such as equalisation and time-alignment can produce improvements in spectral response at high frequencies [11–13]. However, this requires a highly dense dataset of HRTFs capable of artefact free Ambisonic rendering (>2700 to render 35th-order Ambisonics [10]) and is therefore considered impractical and infeasible for individualisation at present, despite techniques such as reciprocity [14] and multiple swept sine [15] offering faster measurement times. Therefore, this paper focuses on virtual loudspeaker rendering of Ambisonic signals, though with pre-encoding of the virtual loudspeaker HRTFs into the SH domain, which allows for dual-band decoding and loudspeaker configurations with more loudspeakers than SH channels both at no added computational cost.

This paper presents a novel method for improving high-frequency reproduction in binaural Ambisonic rendering, without increasing the order of Ambisonics or real-time computational cost. Diffuse-Field Equalisation (DFE) is the removal of the direction-independent aspect of a set of frequency responses measured at all directions on a sphere. By applying the technique of DFE to binaural Ambisonic rendering, high-frequency reproduction can be improved and therefore a more realistic spatial audio experience can be conveyed to the listener.

In a preliminary study, DFE was shown to improve timbre between binaural Ambisonic rendering and standard binaural rendering (HRIR convolution) for 1st-order Ambisonics [16]. This paper expands the investigation to higher order Ambisonics. The methodology of diffuse-field equalisation is explored to determine which quadrature method to use and the optimal number of measurements necessary to produce a sufficient approximation. DFE of Ambisonics is evaluated in three ways; spectral difference over the sphere, predicted sagittal plane localisation and perceptual listening tests. Three orders of Ambisonics: $M = 1$, 3 and 5 are investigated.

2. Ambisonic Rendering

2.1. Encoding

A sound source can be encoded into Ambisonic format for a given location of azimuth θ and elevation ϕ with Ambisonic order M through multiplication with three-dimensional full normalised (N3D) SH functions Y, defined as

$$Y_{mn}^{\sigma}(\theta,\phi) = \sqrt{(2m+1)(2-\delta_{n,0})\frac{(m-n)!}{(m+n)!}}P_{mn}(\sin\phi) \times \begin{cases} \cos(n\theta), \text{ if } \sigma = +1 \\ \sin(n\theta), \text{ if } \sigma = -1 \end{cases}, \quad (1)$$

where $\sigma = \pm 1$, $P_{mn}(\sin\phi)$ are the associated Legendre functions of order m and degree n, and $\delta_{n,0}$ is the Kronecker delta function,

$$\delta_{n,0} \equiv \begin{cases} 1, \text{ for } n = 0 \\ 0, \text{ for } n \neq 0 \end{cases}. \quad (2)$$

The total number of SH channels in an Ambisonic encoded sound field is calculated as $K = (M+1)^2$.

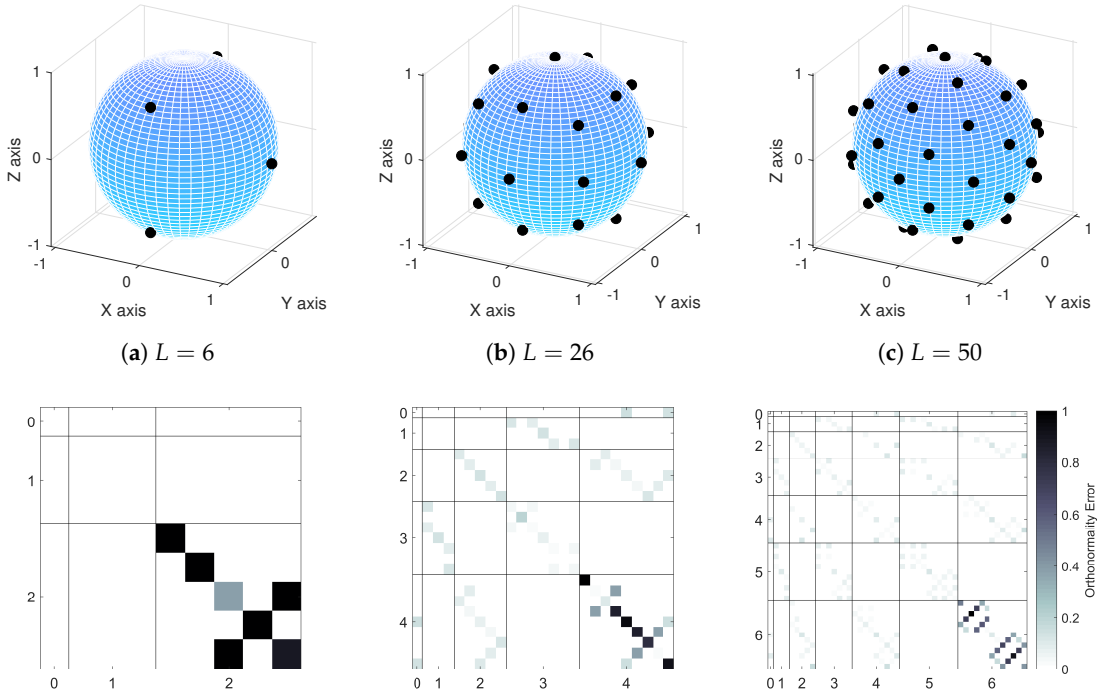
2.2. Decoding

The amount of loudspeakers L required to reproduce the sound field is determined by $L \geq K$, though in the case of $L = K$, which produces the least errors in spatial reproduction, audible timbral artefacts occur when a sound is panned to the exact location of a loudspeaker [17,18]. All loudspeaker configurations used in this paper therefore conformed to the rule $L > K$. The three loudspeaker configurations used in this study for orders $M = 1, 3$ and 5 were Lebedev grid [19] arrangements with

$L = 6$, 26 and 50 loudspeakers respectively. Lebedev grids are quadratures well suited for practical Ambisonic playback due to their near-exact orthonormal capabilities [20] and nesting of the $L = 6$ and $L = 26$ in the $L = 50$ configuration. To test the regularity of the loudspeaker configurations for SH sampling, orthonormality error E_O was calculated as

$$E_O = I_K - \frac{1}{L} C^T \times C, \tag{3}$$

where I_K denotes the $K \times K$ identity matrix, C is the re-encoding matrix and transposition is notated by a superscript ^{T} [18,21]. The layout of loudspeakers and E_O matrix plots of the three Lebedev configurations are presented in Figure 1 for Ambisonic orders up to $M + 1$, which show the regularity for the 6 pt. and near-regularity of the 26 pt. and 50 pt. configurations.



(**a**) $L = 6$ (**b**) $L = 26$ (**c**) $L = 50$

(**d**) E_O for $L = 6$ up to $M = 2$ (**e**) E_O for $L = 26$ up to $M = 4$ (**f**) E_O for $L = 50$ up to $M = 6$

Figure 1. Loudspeaker layouts and orthonormality error matrices for the three Lebedev loudspeaker configurations used in this paper: (**a**) loudspeaker layout for $M = 1$ ($L = 6$), (**b**) loudspeaker layout for $M = 3$ ($L = 26$), (**c**) loudspeaker layout for $M = 5$ ($L = 50$), (**d**) orthonormality error matrix plot for $M = 1$ ($L = 6$), (**e**) orthonormality error matrix plot for $M = 3$ ($L = 26$) and (**f**) orthonormality error matrix plot for $M = 5$ ($L = 50$). In orthonormality error matrix plots, spherical harmonics of different orders are separated to aid visual clarity.

Separate decode matrices were calculated for low and high frequencies for optimal horizontal localisation. Pseudo-inverse mode-matching decoding produces the closest approximation of the original sound field for near-regular loudspeaker arrangements with a non-square re-encoding matrix [22]. For frequencies up to f_{alias} therefore, a pseudo-inverse decoding method with basic channel weighting was used.

Above f_{alias}, where the sound field is inadequately reconstructed, pseudo-inverse decoding with Max $\mathbf{r_E}$ channel weighting was used, which aims to reproduce the energy vector $\mathbf{r_E} = 1$ for all directions [18,22,23]. The gains g_m to be applied to k such to maximise r_E are found from differentiation of r_E with respect to g_m ([18], p. 312), such that

$$\frac{\delta r_E}{\delta g_m} = 0 \Rightarrow r_E(2m+1)g_m = (m+1)g_{m+1} + mg_{m-1}. \quad (4)$$

This recurrence equation can then be rewritten using Legendre polynomials to match Bonnets' Recursion Formula [24], if the rules $g_{-1} = 0$ and $g_{M+1} = 0$ and therefore $g_0 = 1$ and $g_1 = r_E$ are followed, such that $\eta = r_E$ and $g_m = P_m(r_E)$, for orders $m = 0, 1, ..., M$:

$$\eta(2m+1)P_m(\eta) = (m+1)P_{m+1}(\eta) + mP_{m-1}(\eta), \quad (5)$$

where r_E is the largest root of P_{M+1}:

$$P_{M+1} = g_{M+1} = 0. \quad (6)$$

Applying these weights to the Ambisonic channels in the decode process attempts to produce a constant value of the energy vector at all locations on the sphere, which improves high-frequency localisation and interaural level difference (ILD) reproduction [3]. However, in practice Max $\mathbf{r_E}$ weighting reduces the overall amplitude of the decode; a trend that becomes more pronounced as the Ambisonic order increases. This reduction in amplitude is not uniform across the frequency spectrum but rather focused on high frequencies (see Figure 2). The high-frequency attenuation caused by Max $\mathbf{r_E}$ weights was hence negated through amplitude compensation [22,23] prior to the dual-band crossover. The values for each order were calculated from the root-mean-square (RMS) of the Max $\mathbf{r_E}$ channel weights g_m ([18], p. 183). Table 1 presents the RMS Max $\mathbf{r_E}$ channel weights g_{mRMS} for Ambisonic orders 1 to 5.

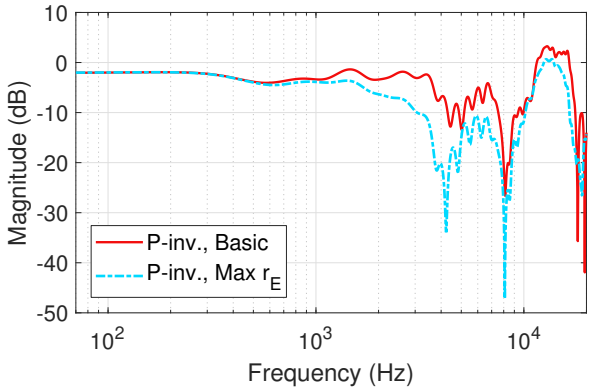


Figure 2. Example of the high-frequency differences of decoding with and without Max $\mathbf{r_E}$ weighting-fifth-order binaural Ambisonic rendering (Neumann KU 100 HRIRs) at (θ°, ϕ°) = (0, 0) (left ear).

Table 1. RMS Max $\mathbf{r_E}$ weightings for Ambisonic orders 1 to 5.

M	**1**	**2**	**3**	**4**	**5**
g_{mRMS}	0.707	0.633	0.600	0.581	0.569

Binaural decoders were computed by encoding the HRTFs of each loudspeaker configuration into a continuous SH representation by a summation of the multiplication of each channel of the decoding matrix with the corresponding HRTF such that

$$D_K^{SH} = \sum_{l=1}^{L} H_l \times D_l, \quad (7)$$

where H is the HRTF and D is the decode matrix. This was repeated for both Basic and Max $\mathbf{r_E}$ weighted decoders. HRIRs in this study were from the Neumann KU 100 dummy head [25].

Crossover filters between the low- and-high frequency decoders were phase matched to avoid unwanted destructive interference around the crossover frequency [22,26]. Whereas in loudspeaker reproduction, the transition between low and high frequencies should be implemented gradually [26,27], when rendering Ambisonics binaurally the crossover can be implemented more abruptly and at a higher frequency due to the head's fixed position inside the virtual loudspeaker array. Crossover frequencies were calculated from the integrated D-error, which is the error between a wave field constructed from a single plane wave and an Ambisonically reconstructed plane wave [18,28,29]. For Ambisonic orders 1 to 5, dual-band decoder crossover frequencies are presented in Table 2.

Table 2. Dual-band decode crossover frequencies producing 20% integrated D-error for Ambisonic orders 1 to 5, according to [23].

M	**1**	**2**	**3**	**4**	**5**
f (Hz)	743	1346	1960	2595	3230

The two binaural decoders were high and low-passed using the above filters and combined. An advantage of using an SH HRTF format binaural decoder is that the decoding matrix takes length K instead of L, which removes the increased computation of additional convolutions in the case $L > K$.

Binaural rendering A is then achieved by convolution of the encoded Ambisonic format signals with the binaural decoder:

$$A = \sum_{k=1}^{K} B_k * D_k^{SH}, \tag{8}$$

where B_K is the encoded Ambisonic signal, D_K^{SH} is the dual-band binaural decoder and $*$ denotes convolution.

3. Diffuse-Field Equalisation

3.1. Diffuse-Field Response Calculation

The diffuse-field response of the SH binaural Ambisonic decoder D_k^{SH} can be calculated from a sum of the RMS of the SH channels, a process referred to as numerical integration. However, in this study an alternative approach was taken through spatial sampling of the sphere (which was found to give equivalent results), so that further developments could be explored, such as directional bias in the diffuse-field response for localised spectral improvements [30].

The Ambisonic diffuse-field response A_{diff}, calculated from the RMS of Q measurements, was approximated (separately for each ear) by:

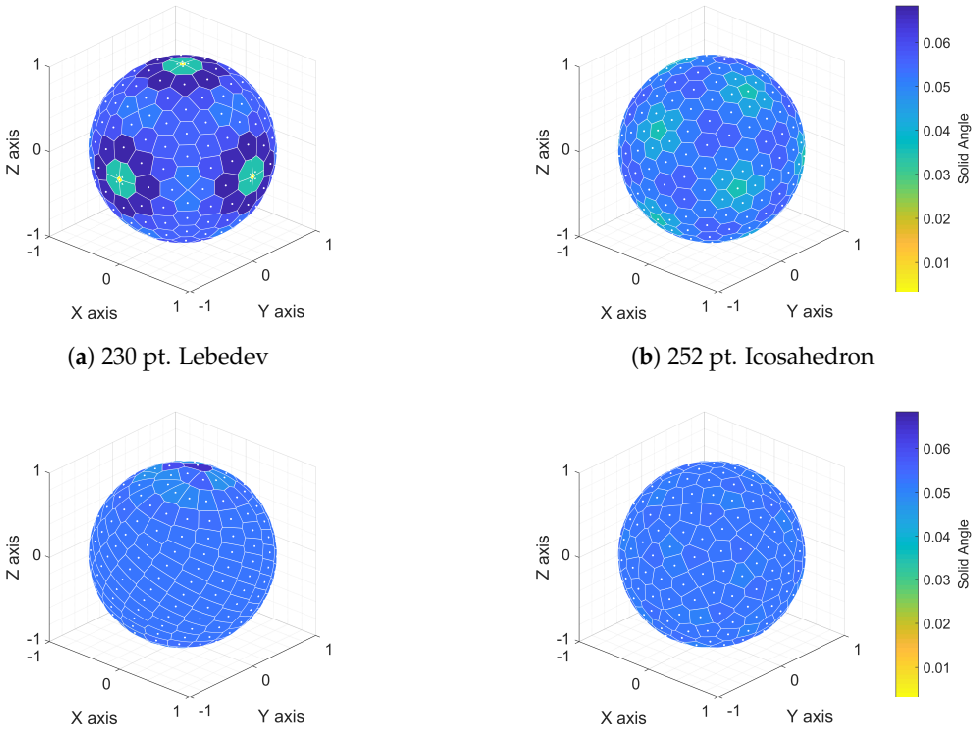
$$A_{\text{diff}} = \sqrt{\frac{1}{Q}\sum_{q=1}^{Q} \Omega_q {A_q}^2}, \tag{9}$$

where A_q is a binaural Ambisonic render and Ω_q is the relative solid angle weight of the virtual sound source position, which is calculated as the area subtended by its position on the sphere.

In order to determine the optimal quadrature method and number of points necessary to produce an adequate approximate diffuse-field response, four quadrature methods were investigated with varying number of points by rendering approximate diffuse-field responses.

Four quadrature methods were investigated for the distribution of points on a sphere: the Lebedev grid [19], Icosahedron division [31], Fibonacci spiral [32] and spherical T-design [33]. Voronoi sphere plots of the quadrature methods with a similar number of points are shown in Figure 3 to compare the regularity of the quadrature methods. The plots show T-design quadrature produces the highest regularity of the four methods. Simulated diffuse-field responses using the four quadrature methods

differed by up to ± 1.5 dB at 1 kHz, however implementation of solid angle weighting brought the variation to below ± 0.1 dB. Therefore, providing solid angle weighting is implemented, the quadrature method need not be highly regular.



(**a**) 230 pt. Lebedev (**b**) 252 pt. Icosahedron

(**c**) 240 pt. Fibonacci (**d**) 240 pt. 21-Design

Figure 3. Voronoi sphere plots demonstrating the regularity in spherical distribution of points for four quadrature methods: (**a**) 230 pt. Lebedev, (**b**) 252 pt. Icosahedron, (**c**) 240 pt. Fibonacci and (**d**) 240 pt. 21-Design.

The minimum number of measurements necessary to calculate a sufficient approximation of the diffuse-field was investigated by rendering diffuse-field responses with a varying number of measurements. The number of measurements ranged from $Q = L$ to $Q = 1 \times 10^5$ using Fibonacci quadrature, for the three tested orders of Ambisonics, and validated through comparison to the numerical integration method. With solid angle weighting implementation, variation in calculated diffuse-field response was as little as $\pm 1 \times 10^{-4}$ dB at 20 kHz when $Q > 4L$, which will be perceptually negligible. Therefore 240 pt. T-design quadrature was used for the remainder of the study to ensure minimal error between the numerical integration method.

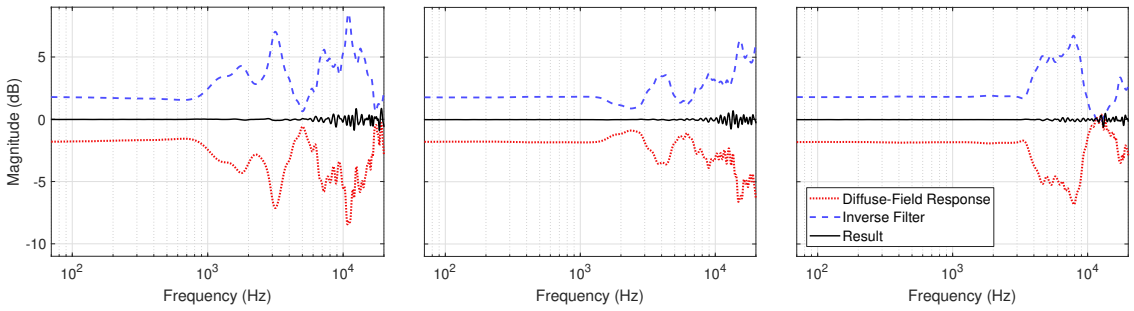
3.2. Inverse Filter Calculation

For the three tested orders of Ambisonics, 240 points around the sphere, distributed using T-design quadrature (21-design), were encoded into binaural Ambisonics. Diffuse-field responses were simulated from the RMS of each binaural Ambisonic render with solid-angle weighting separately for each ear.

Linear-phase inverse filters were then calculated from the diffuse-field responses using Kirkeby and Nelson's least-mean-square regularisation method [34], which produces perceptually preferred inversions to other currently available methods [35]. 1/4 octave smoothing was implemented using the complex smoothing approach of [36], and the range of inversion was 2 Hz–20 kHz, with in-band and out-band regularisation of 25 dB and 5 dB, respectively. The target response of the inverse filter was the diffuse-field response of the original HRTF dataset.

The diffuse-field responses, inverse filters and resulting equalised frequency responses of the 1st, 3rd and 5th-order binaural Ambisonic loudspeaker configurations are presented in Figure 4. The plots

show how the diffuse-field responses vary significantly for all 3 orders above f_{alias}, with deviations as large as 9 dB for 1st order.



(**a**) $M = 1$ (**b**) $M = 3$ (**c**) $M = 5$

Figure 4. Diffuse-field response, inverse filters and resulting responses of the three tested Ambisonic configurations (left ear): (**a**) $M = 1$, (**b**) $M = 3$ and (**c**) $M = 5$.

DFE was implemented through offline convolution of the calculated inverse filters with the HRIRs used in binaural Ambisonic rendering for each loudspeaker configuration. With truncation of the processed HRIRs, implementation therefore comes at no additional real-time computational cost.

To assess whether the need for DFE applies to other HRIR datasets, the diffuse-field response calculations were repeated for 5th-order using the 18 human datasets of individualised HRIRs from the SADIE II (Spatial Audio for Domestic Interactive Entertainment: https://www.york.ac.uk/sadie-project/.) database [37]. The calculated diffuse-field responses are illustrated in Figure 5, which shows the necessity for diffuse-field equalisation for all HRIR datasets.

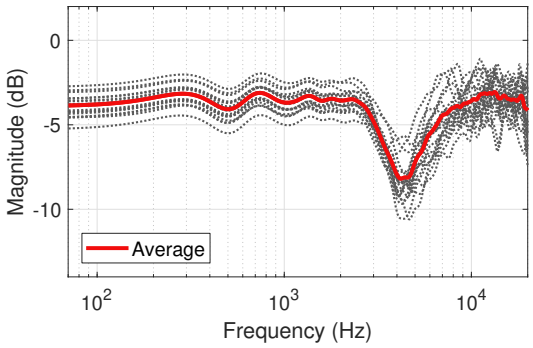


Figure 5. Diffuse-field responses of the 50 pt. Lebedev loudspeaker configuration for $M = 5$ for the 18 human subjects of the SADIE II database, with average response (left ear).

4. Evaluation

DFE of Ambisonics was evaluated in three ways by comparing binaural Ambisonic renders, with and without DFE, to a reference HRIR data set [25] of 16020 HRIRs (89 elevations at 180 azimuths in 2° increments). The objective change in spectral difference between binaural Ambisonic rendering and standard binaural rendering (HRIR convolution) when implementing DFE was investigated, to measure the influence of DFE over all directions on the sphere. Sagittal plane localisation was assessed through a binaural model, which allows evaluation of many more directions than would be feasible in a listening test. Finally, two listening tests were conducted to measure the perceptual effect of DFE on timbral similarity between binaural Ambisonic rendering and standard binaural rendering.

For each angle of the reference data set, binaural Ambisonic renders were generated and diffuse-field equalised through convolution of each render with the inverse filters of that order of

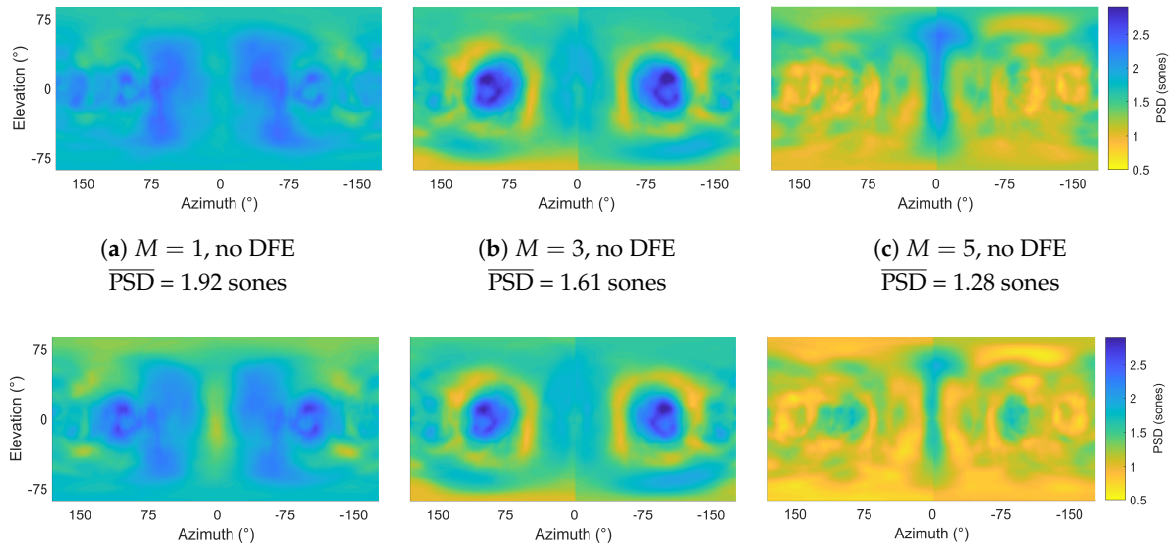
Ambisonics. The reference, Ambisonic and DFE Ambisonic HRIRs were truncated to 1024 taps with 50 sample in/out half-Hanning windows applied. Linear phase was preserved.

4.1. Spectral Difference

Basic spectral difference, calculated from the difference between the magnitude value of each frequency bin of the fast Fourier transforms (FFTs) of two audio files, is not an accurate metric for human auditory perception on its own, as the human auditory system's sensitivity differs greatly depending on relative amplitude, frequency and temporal aspects [38]. Therefore, the objective spectral difference calculation aimed to more accurately represent the perceptual effect of absolute spectral differences, and is here on in referred to as perceptual spectral difference (PSD) [39].

PSD was calculated between the Ambisonic HRIRs and reference HRIRs as follows. The HRIR data sets were converted to HRTFs, the frequency domain equivalent of HRIRs, using an FFT with a window size of 4096 samples. Each frequency bin value was amplitude weighted according to ISO Standard 226 equal loudness contours [40] to account for human frequency-dependent amplitude sensitivity. The value of each frequency bin was then weighted to reflect human perception of loudness using the sone scale, at a ratio of +10 phons per doubling of perceived loudness [38,41,42], which also accounts for human auditory features such as spectral peaks being more perceptually significant than notches [43]. Equivalent rectangular bandwidth (ERB) filters were utilised to compensate for the linear frequency interval sampling of the FFT. Rendered data sets were normalised to the mean perceptual loudness level of the reference data set, and PSD was calculated as the absolute difference between each frequency bin of the two data sets.

PSD between reference HRIRs and Ambisonically generated HRIRs, with and without DFE, was calculated for 16020 directions for each tested order. Figure 6 displays the mean absolute values of PSD between Ambisonic HRIRs and reference HRIRs with and without DFE over the sphere (mean of left and right ear PSD calculations). The figures show DFE implementation improves spectral reproduction for a large amount of the sphere, particularly for 1st and 5th-orders, but makes it worse at lateral directions.



(**a**) $M = 1$, no DFE
$\overline{\text{PSD}}$ = 1.92 sones

(**b**) $M = 3$, no DFE
$\overline{\text{PSD}}$ = 1.61 sones

(**c**) $M = 5$, no DFE
$\overline{\text{PSD}}$ = 1.28 sones

(**d**) $M = 1$, with DFE
$\overline{\text{PSD}}$ = 1.80 sones

(**e**) $M = 3$, with DFE
$\overline{\text{PSD}}$ = 1.58 sones

(**f**) $M = 5$, with DFE
$\overline{\text{PSD}}$ = 1.11 sones

Figure 6. $\overline{\text{PSD}}$ between HRTFs and reference HRTFs with and without DFE (mean of left and right ear PSD calculations): (**a**) $M = 1$, no DFE, (**b**) $M = 3$, no DFE, (**c**) $M = 5$, no DFE, (**d**) $M = 1$, with DFE, (**e**) $M = 3$, with DFE and (**f**) $M = 5$, with DFE. PSD: Perceptual spectral difference; HRTF: Head-related transfer function; DFE: Diffuse-field equalisation.

The mean absolute values of PSD between reference HRIRs and Ambisonic HRIRs with and without DFE across all angles on the sphere are presented in Table 3. Values were solid-angle weighted to account for the clustering of points at the poles in Gaussian quadrature. The table shows that DFE reduces the overall PSD between Ambisonic HRTFs and reference HRTFs for all three tested configurations.

Table 3. Solid angle weighted mean values of PSD between reference HRTFs and Ambisonically rendered HRTFs for left and right ears.

M	**1**		**3**		**5**	
DFE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
$\overline{\text{PSD}}$ (sones)	1.92	1.80	1.61	1.58	1.28	1.11

4.2. Sagittal Plane Localisation

The effect of DFE on elevation localisation in the sagittal plane was assessed using a perceptual model [44], which compares two data sets of HRIRs (in this case Ambisonic and reference), and simulates human auditory processing to predict height localisation. It produces two psychoacoustic performance metrics: quadrant error (QE), a prediction of localisation confusion (presented as a percentage), and polar RMS error (PE), a prediction of precision and accuracy in degrees. The frequency range of the model's filterbank was set to 1.5–18 kHz due to the KU 100's lack of torso and therefore no elevation cues below 1.5 kHz [45], and the upper limit of human hearing.

Table 4 shows the predicted QE and PE values for Ambisonic orders 1, 3 and 5 and Figure 7 illustrates predicted sagittal plane localisation. Results indicate improved localisation performance with the implementation of DFE for all three tested configurations.

Table 4. Performance Predictions of binaural Ambisonic rendering using the Sagittal Plane Localisation Model [44], with and without DFE.

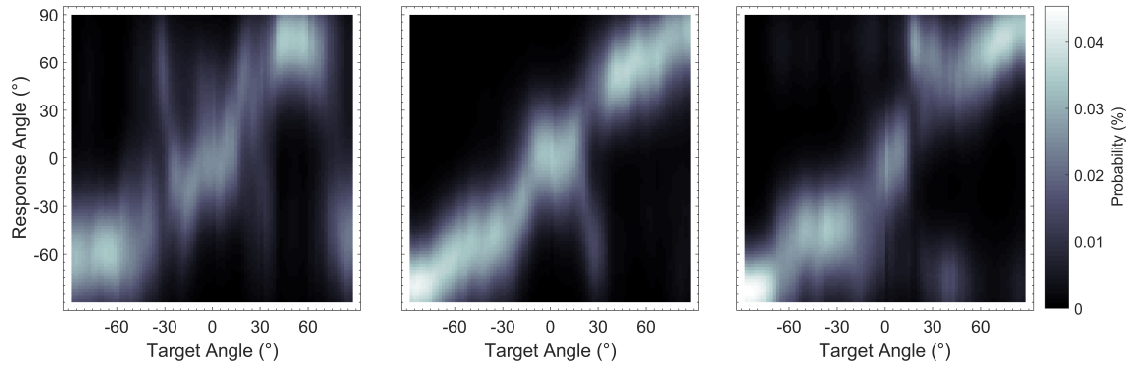
M	**1**		**3**		**5**	
DFE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
QE (%)	11.9	9.8	3.6	2.8	9.8	6.2
PE (°)	35.8	35.2	27.6	26.6	30.6	27.7

4.3. Perceptual Listening Tests

To evaluate the perceptual effect of DFE on timbre in binaural Ambisonic rendering, two listening tests were conducted on 20 participants aged between 20 to 38 (17 male, 2 non-binary, 1 female) with normal hearing (ISO Standard 389 [46]) and prior critical listening experience (such as education or employment in audio or music engineering).

Tests were conducted in a quiet room using a single set of Sennheiser HD 650 circum-aural headphones and an Apple MacBook Pro with a Fireface UCX audio interface, which has software controlled input and output levels. The headphones were equalised using a Neumann KU 100 dummy head from 11 measurements using the swept sine impulse response technique [47] with re-fitting of the headphones between measurements and 1 octave band smoothing in the inverse filter.

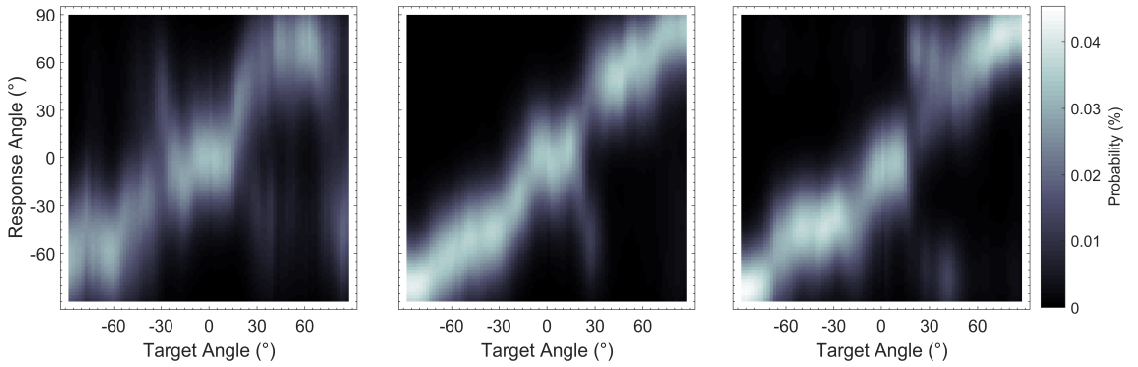
Prior to the tests, participants were given the ANSI S1.1-1994 definition of timbre [48] and taken through a training exercise to familiarise themselves with the graphical user interface and task.



(**a**) $M = 1$, no DFE
QE = 11.9%, PE = 35.8°

(**b**) $M = 3$, no DFE
QE = 3.6%, PE = 27.6°

(**c**) $M = 5$, no DFE
QE = 9.8%, PE = 30.6°



(**d**) $M = 1$, with DFE
QE = 9.8%, PE = 35.2°

(**e**) $M = 3$, with DFE
QE = 2.8%, PE = 26.6°

(**f**) $M = 5$, with DFE
QE = 6.2%, PE = 27.7°

Figure 7. Sagittal Plane Localisation Model [44] plots of binaural Ambisonic rendering, with and without Diffuse-Field Equalisation (DFE): (**a**) $M = 1$, no DFE, (**b**) $M = 3$, no DFE, (**c**) $M = 5$, no DFE, (**d**) $M = 1$, with DFE, (**e**) $M = 3$, with DFE and (**f**) $M = 5$, with DFE. Plots show greater clarity (bolder shading) with DFE implementation.

The base stimulus was one second of monophonic pink noise at a sample rate of 48 kHz, windowed by onset and offset half-Hanning ramps of 5 ms. Each test sound was generated by convolving the pink noise with a HRIR; either Ambisonic or not. The test sound locations (ψ) corresponded to the central points of the faces of a dodecahedron. To reduce the total number of trials, symmetry was assumed and thus only locations in the left hemisphere were used, amounting to 8 locations (see Table 5). Test sounds were normalised to a corresponding A-weighted level of between 65 and 70 dB SPL which is in line with conversational speech listening levels.

Table 5. Spherical coordinates of test sound locations.

ψ	1	2	3	4	5	6	7	8
θ (°)	180	50	118	0	180	62	130	0
ϕ (°)	64	46	16	0	0	−16	−46	−64

4.3.1. Test Paradigms

The first listening test followed the multiple stimulus test with hidden reference and anchor (MUSHRA) paradigm, ITU-R BS.1534-3 [49]. The reference was a direct HRIR convolution, and medium and low anchors were low-pass filtered versions of the reference stimulus with an f_c of 7 kHz and 3.5 kHz, respectively. The other 6 stimuli were the binaural Ambisonic renders for three Ambisonic

orders, with and without DFE. For each trial, the listener was asked to rate the 9 stimuli in terms of timbral similarity to the reference. The 8 test sound locations were repeated giving a total of 16 trials. The presentation of stimuli and trials was randomised and double blind.

The second listening test was an AB comparison. Participants were presented with two sets of three consecutive stimuli (corresponding to Ambisonic renders of 1st, 3rd and 5th-orders), one set of which was diffuse-field equalised, and were asked to rate them in terms of timbral consistency. The 8 test sound locations were repeated with a different arrangement of the Ambisonic orders (the first was 1, 3, 5 and the second was 1, 5, 3), giving a total of 16 trials. The presentation of trials was randomised and double blind.

4.3.2. Results

No participants results were excluded, based on the criteria of rating the hidden reference less than 90% for more than 15% of trials or rating the mid-anchor higher than 90% for more than 15% of trials. The results from both listening tests were tested for normality using the Kolmogorov-Smirnov test, which showed all data as non-normally distributed. As a result, all statistical analysis was conducted using non-parametric methods.

The median results of the MUSHRA test, conducted to determine whether DFE reduces the differences in timbre between binaural Ambisonic rendering and HRIR convolution, are shown in Figure 8 for each condition across all test sound locations, with non-parametric 95% confidence intervals (CI) [50].

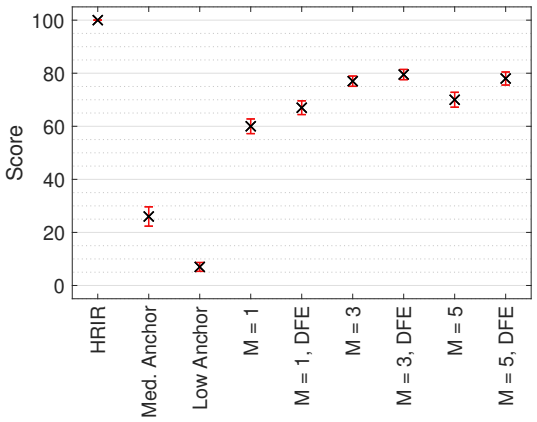
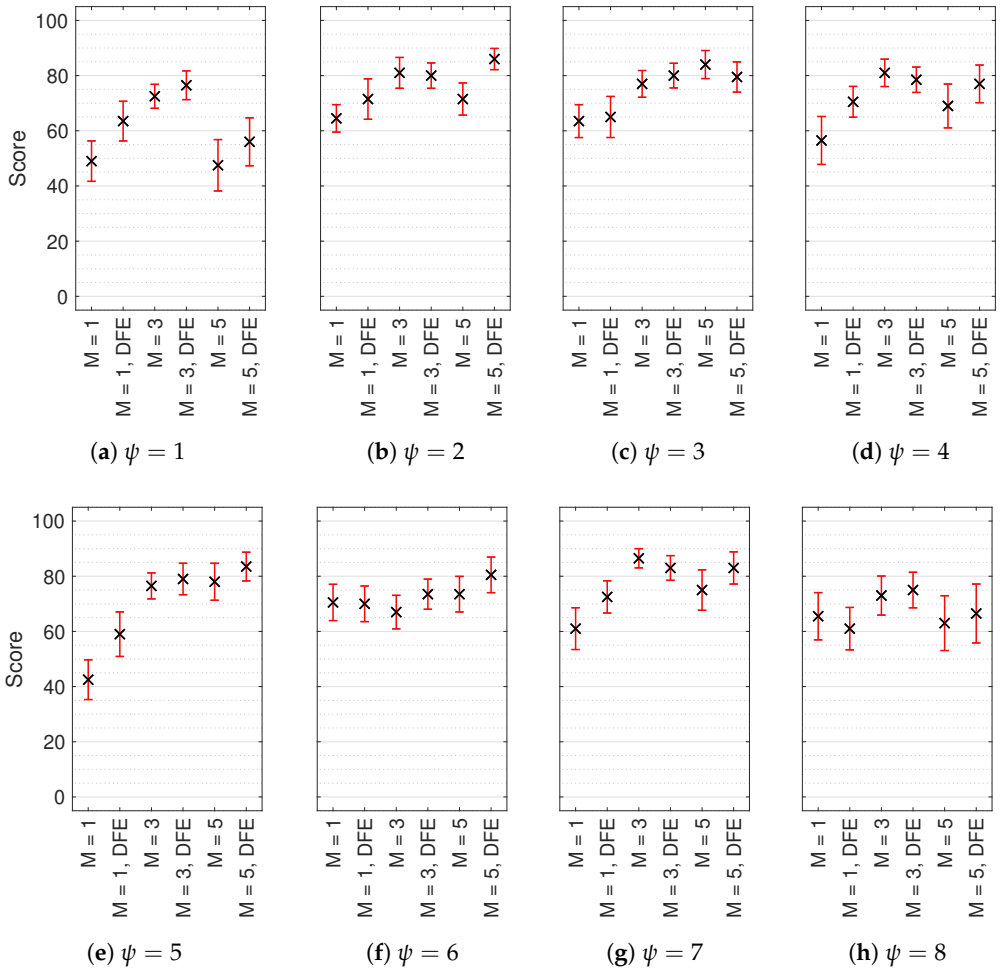


Figure 8. Median MUSHRA results with non-parametric 95% CI across all test sound locations. Score indicates perceived timbral similarity between test stimulus and HRIR reference.

Friedman's Analysis of Variance (ANOVA) tests showed a statistically significant difference ($\chi^2(5) = 247.6, p < 0.05$) between standard and DFE binaural Ambisonic rendering for all tested orders and sound locations. 1st-order Ambisonics showed the most improvement, followed by 5th and 3rd.

The perceptual effect of DFE was found to vary with test sound location, with a Friedman's ANOVA showing this variation to be statistically significant ($\chi^2(7) = 127.8, p < 0.05$). Figure 9 shows the median results with non-parametric 95% CI for each test sound location ψ.

Post-hoc Wilcoxon signed-rank tests determined which test conditions with DFE produced a significant improvement in timbre; the results of which are shown in Table 6. For both 1st and 5th-order Ambisonics, DFE was shown to bring the timbre of binaural Ambisonic rendering closer to HRIR convolution with statistical significance for 5 of the 8 test sound locations. Results for 3rd-order were much less clear and only significant for one test sound location.



(**a**) $\psi = 1$ (**b**) $\psi = 2$ (**c**) $\psi = 3$ (**d**) $\psi = 4$

(**e**) $\psi = 5$ (**f**) $\psi = 6$ (**g**) $\psi = 7$ (**h**) $\psi = 8$

Figure 9. Median multiple stimulus test with hidden reference and anchor (MUSHRA) results with non-parametric 95% confidence intervals (CI) for each test sound location (ψ): (**a**) $\psi = 1$, (**b**) $\psi = 2$, (**c**) $\psi = 3$, (**d**) $\psi = 4$, (**e**) $\psi = 5$, (**f**) $\psi = 6$, (**g**) $\psi = 7$ and (**h**) $\psi = 8$. Reference and anchor scores omitted. Score indicates perceived timbral similarity between test stimulus and HRIR reference.

Table 6. Hypothesis test results of the MUSHRA test results of the three Ambisonic orders for each test sound location using Wilcoxon signed-rank test (1 indicates statistical significance at $p < 0.05$; * indicates $p < 0.01$).

ψ	1	2	3	4	5	6	7	8
h (M = 1)	1	1 *	0	1 *	1 *	0	1 *	0
h (M = 3)	0	0	0	0	0	1 *	0	0
h (M = 5)	1 *	1 *	0	0	1	1 *	1 *	0

The median results of the second listening test, the AB comparison conducted to determine whether DFE improved the consistency of timbre between Ambisonic orders, are shown in Figure 10 for both conditions across all test sound locations, with non-parametric 95% CI.

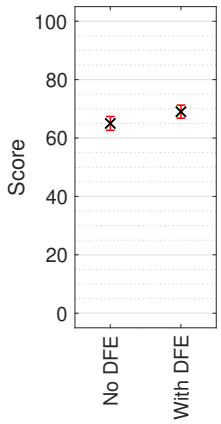
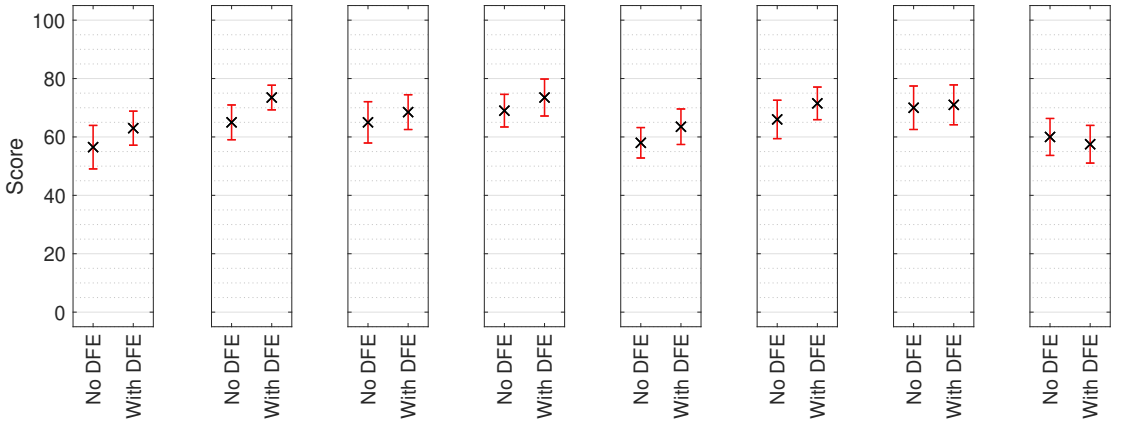


Figure 10. Median AB results with non-parametric 95% CI across all test sound locations. Score indicates perceived timbral consistency between the three tested orders of Ambisonics.

Overall across all test sound locations, DFE produced higher timbral consistency between different Ambisonic orders, and a Friedman's ANOVA test showed that this was statistically significant ($\chi^2(1) = 8.45, p < 0.05$). To assess how perceived timbral consistency varied with test sound location, a second Friedman's ANOVA was conducted and showed significance ($\chi^2(7) = 37.5, p < 0.05$). Figure 11 shows the median AB results with non-parametric 95% CI for each test sound location ψ. Post-hoc Wilcoxon signed-rank tests to determine which test sound locations produced statistically significant results were conducted; the results of which are displayed in Table 7.



(**a**) $\psi = 1$ (**b**) $\psi = 2$ (**c**) $\psi = 3$ (**d**) $\psi = 4$ (**e**) $\psi = 5$ (**f**) $\psi = 6$ (**g**) $\psi = 7$ (**h**) $\psi = 8$

Figure 11. Median AB results with non-parametric 95% CI for each test sound location (ψ): (**a**) $\psi = 1$, (**b**) $\psi = 2$, (**c**) $\psi = 3$, (**d**) $\psi = 4$, (**e**) $\psi = 5$, (**f**) $\psi = 6$, (**g**) $\psi = 7$ and (**h**) $\psi = 8$. Reference and anchor scores omitted. Score indicates perceived timbral consistency between the three tested orders of Ambisonics.

Table 7. Hypothesis tests of the AB test results for each test sound location using Wilcoxon signed-rank test (1 indicates statistical significance at $p < 0.05$; * indicates $p < 0.01$).

ψ	1	2	3	4	5	6	7	8
h	0	1 *	0	1	0	0	0	0

5. Discussion

The results showed that binaural Ambisonic rendering with DFE produced an overall improvement in high-frequency reproduction over standard binaural Ambisonic rendering when compared to direct HRIR rendering, as well as improved timbral consistency between different orders of Ambisonics. However, results were not hugely significant, indicating that even with DFE, binaural Ambisonic reproduction still varies considerably in timbre between orders and between direct HRIR rendering.

Results for the 1st and 5th-order configurations were more substantial than 3rd-order, which is likely due to the greater variation in diffuse-field responses for the 1st and 5th-order configurations, and at frequencies with more perceptual importance (see again Figure 4). Thus for these two configurations, equalisation produced a more significant improvement.

Results were shown to vary with sound source location, something that was evident in both the perceptual spectral difference calculations across the sphere and the results of the listening tests. However, no trends between the influence of DFE and specific sound source location attributes (such as increasing elevation or increasing lateralisation) were found. A possible general explanation for the directional variation in influence of DFE is that different sound source locations produce HRTFs that emphasise or cut different frequencies, and a boost or cut to a specific frequency band (i.e. from the diffuse-field equalisation) will therefore affect certain locations more than others.

6. Conclusions

The inaccuracies of high-frequency reproduction in Ambisonics, caused by comb filtering from the summation of multiple analogous signals at the ears, have been addressed in this paper through the application of the diffuse-field equalisation technique. By implementing diffuse-field equalisation in binaural Ambisonic rendering as a low-computation additional stage to the binaural rendering process, the diffuse-field response of the binaural Ambisonic loudspeaker configuration is flattened out, which changes the frequency response of renders at individual sound source locations. This has been shown to, on average over all directions on the sphere, bring the spectral reproduction of binaural Ambisonic rendering closer to direct HRIR rendering.

For the three tested loudspeaker configurations, corresponding to 1st, 3rd and 5th-order Ambisonics, DFE produced small improvements in spectral difference over the sphere and sagittal plane localisation predictions. Listening tests on timbre corroborated this, though not all test conditions were statistically significant. Listening tests also showed that DFE produces a small improvement in timbral consistency between different orders of Ambisonics.

This paper has shown that a low-computation equalisation pre-processing stage can produce an incremental improvement in the high-frequency reproduction of binaural Ambisonic rendering, however there still exists a significant difference between binaural Ambisonic rendering using the virtual loudspeaker approach and direct HRIR rendering. Therefore, diffuse-field equalisation alone is not enough to eradicate the timbral issues posed by Ambisonic rendering, and should be used as a basis for future research in improving Ambisonic rendering at high frequencies.

Author Contributions: Conceptualisation, T.M. and G.K.; Methodology, T.M.; Software, T.M.; Validation, T.M., D.T.M. and G.K.; Formal Analysis, T.M.; Investigation, T.M.; Resources, T.M.; Data Curation, T.M.; Writing—Original Draft Preparation, T.M.; Writing—Review & Editing, T.M., D.T.M. and G.K.; Visualisation, T.M.; Supervision, D.T.M. and G.K.; Project Administration, T.M.; Funding Acquisition, G.K.

Funding: This research was funded by a Google Faculty Research Award and the Engineering and Physical Sciences Research Council (EP/M001210/1).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DFE	Diffuse-field equalisation
HRIR	Head-related transfer function
SH	Spherical harmonic
N3D	Three-dimensional full normalisation
ILD	Interaural level difference
RMS	Root-mean-square
FFT	Fast Fourier transform
PSD	Perceptual spectral difference
HRTF	Head-related transfer function
ERB	Equivalent rectangular bandwidth
QE	Quadrant error
PE	Polar RMS error
MUSHRA	Multiple stimulus test with hidden reference and anchor
CI	Confidence intervals
ANOVA	Analysis of variance

References

1. Gerzon, M.A. Periphony: With-Height Sound Reproduction. *J. Audio Eng. Soc.* **1973**, *21*, 2–10.
2. Gerzon, M.A. Criteria For Evaluating Surround-Sound Systems. *J. Audio Eng. Soc.* **1977**, *25*, 400–408.
3. Gerzon, M. General Metatheory of Auditory Localization. In Proceedings of the 92nd Convention of the Audio Engineering Society, Vienna, Austria, 24–27 March 1992.
4. Bregman, A.S. *Auditory Scene Analysis: The Perceptual Organization of Sound*; The MIT Press: Cambridge, MA, USA, 1990.
5. Rumsey, F.; Zieliński, S.; Kassier, R.; Bech, S. On the Relative Importance of Spatial and Timbral Fidelities in Judgments of Degraded Multichannel Audio Quality. *J. Acoust. Soc. Am.* **2005**, *118*, 968–976. [CrossRef] [PubMed]
6. Jot, J.M.; Wardle, S.; Larcher, V. Approaches to Binaural Synthesis. In Proceedings of the 105th Convention of the Audio Engineering Society, San Francisco, CA, USA, 26–29 September 1998.
7. Noisternig, M.; Sontacchi, A.; Musil, T.; Höldrich, R. A 3D Ambisonic Based Binaural Sound Reproduction System. In Proceedings of the AES 24th International Conference on Multichannel Audio, Banff, AB, Canada, 26–28 June 2003.
8. Rafaely, B.; Avni, A. Interaural Cross Correlation in a Sound Rield Represented by Spherical Harmonics. *J. Acoust. Soc. Am.* **2010**, *127*, 823–828. [CrossRef] [PubMed]
9. Avni, A.; Ahrens, J.; Geier, M.; Spors, S.; Wierstorf, H.; Rafaely, B. Spatial Oerception of Sound Fields Recorded by Spherical Microphone Arrays with Varying Spatial Resolution. *J. Acoust. Soc. Am.* **2013**, *133*, 2711–2721. [CrossRef] [PubMed]
10. Bernschütz, B.; Vázquez Giner, A.; Pörschmann, C.; Arend, J. Binaural Reproduction of Plane Waves with Reduced Modal Order. *Acta Acust. United Acust.* **2014**, *100*, 972–983. [CrossRef]
11. Ben-Hur, Z.; Brinkmann, F.; Sheaffer, J.; Weinzierl, S.; Rafaely, B. Spectral Equalization in Binaural Signals Represented by Order-Rruncated Spherical Harmonics. *J. Acoust. Soc. Am.* **2017**, *141*, 4087–4096. [CrossRef] [PubMed]
12. Zaunschirm, M.; Schoerkhuber, C.; Hoeldrich, R. Binaural Rendering of Ambisonic Signals by HRIR Time Alignment and a Diffuseness Constraint. *J. Acoust. Soc. Am.* **2018**, *3616*. [CrossRef]
13. Schörkhuber, C.; Zaunschirm, M.; Höldrich, R. Binaural Rendering of Ambisonic Signals via Magnitude Least Squares. In Proceedings of the DAGA 2018: 44. Deutsche Jahrestagung für Akustik, Munich, Germany, 19–22 March 2018; pp. 339–342.
14. Zotkin, D.N.; Duraiswami, R.; Grassi, E.; Gumerov, N.A. Fast Head-Related Transfer Function Measurement via Reciprocity. *J. Acoust. Soc. Am.* **2006**, *120*, 2202–2215. [CrossRef] [PubMed]
15. Majdak, P.; Balazs, P.; Laback, B. Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. *J. Audio Eng. Soc.* **2007**, *55*, 623–636.

16. McKenzie, T.; Murphy, D.; Kearney, G. Diffuse-Field Equalisation of First-Order Ambisonics. In Proceedings of the 20th International Conference on Digital Audio Effects (DAFx), Edinburgh, UK, 5–9 September 2017; pp. 389–396.
17. Poletti, M.A. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. *J. Audio Eng. Soc.* **2005**, *53*, 1004–1024.
18. Daniel, J. Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia. Ph.D. Thesis, l'Université Paris, Paris, France, 2000.
19. Lebedev, V.I. Quadratures on a Sphere. *J. USSR Comput. Math. Math. Phys.* **1976**, *16*, 10–24. [CrossRef]
20. Lecomte, P.; Gauthier, P.A.; Langrenne, C.; Berry, A.; Garcia, A. A Fifty-Node Lebedev Grid and its Applications to Ambisonics. *J. Audio Eng. Soc.* **2016**, *64*, 868–881. [CrossRef]
21. Moreau, S.; Daniel, J.; Bertet, S. 3D Sound Field Recording With Higher Order Ambisonics-Objective Measurements and Validation of Spherical Microphone. In Proceedings of the 120th Convention of the Audio Engineering Society, Paris, France, 20–23 May 2006; pp. 1–24.
22. Gerzon, M.A.; Barton, G.J. Ambisonic Decoders for HDTV. In Proceedings of the 92nd Convention of the Audio Engineering Society, Vienna, Austria, 24–27 March 1992.
23. Daniel, J.; Rault, J.B.; Polack, J.D. Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. In Proceedings of the 105th Convention of the Audio Engineering Society, San Francisco, CA, USA, 26–29 September 1998.
24. Morse, P.M.; Ingard, U. *Theoretical Acoustics*; Princeton University Press: Princeton, NJ, USA, 1968.
25. Bernschütz, B. A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100. In Proceedings of the AIA-DAGA Conference on Acoustics, Merano, Italy, 18–21 March 2013; pp. 592–595.
26. Heller, A.J.; Lee, R.; Benjamin, E.M. Is My Decoder Ambisonic? In Proceedings of the 125th Convention of the Audio Engineering Society, San Francisco, CA, USA, 2–5 October 2008; Convention Paper 7553.
27. Farina, A. Software Implementation of B-Format Encoding and Decoding. In Proceedings of the 104th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 16–19 May 1998.
28. Bamford, J.S.; Vanderkooy, J. Ambisonic Sound for Us. In Proceedings of the 99th Convention of the Audio Engineering Society, New York, NY, USA, 6–9 October 1995.
29. Poletti, M. The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems. *J. Audio Eng. Soc.* **1996**, *44*, 948–963.
30. McKenzie, T.; Murphy, D.; Kearney, G. Directional Bias Equalisation of First-Order Binaural Ambisonic Rendering. In Proceedings of the AES Conference on Audio for Virtual and Augmented Reality, Redmond, WA, USA, 20–22 August 2018. [CrossRef]
31. Burkardt, J. SPHERE_GRID—Points, Lines, Faces on a Sphere. Available online: http://people.sc.fsu.edu/~jburkardt/datasets/sphere_grid/sphere_grid.html (accessed on 15 September 2018).
32. Saff, E.B.; Kuijlaars, A.B.J. Distributing Many Points on a Sphere. *J. Math. Intell.* **1997**, *19*, 5–11. [CrossRef]
33. Hardin, R.H.; Sloane, N.J.A. McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions. *J. Disc. Comput. Geometry* **1996**, *15*, 429–441. [CrossRef]
34. Kirkeby, O.; Nelson, P.A. Digital Filter Design for Inversion Problems in Sound Reproduction. *J. Audio Eng. Soc.* **1999**, *47*, 583–595.
35. Schärer, Z.; Lindau, A. Evaluation of Equalization Methods for Binaural Signals. In Proceedings of the 126th Convention of the Audio Engineering Society, Munich, Germany, 7–10 May 2009.
36. Hatziantoniou, P.D.; Mourjopoulos, J.N. Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses. *J. Audio Eng. Soc.* **2000**, *48*, 259–280.
37. Kearney, G.; Doyle, T. A HRTF Database for Virtual Loudspeaker Rendering. In Proceedings of the 139th Convention of the Audio Engineering Society, New York, NY, USA, 29 October–1 November 2015.
38. Wang, S.; Sekey, A.; Gersho, A. An Objective Measure for Predicting Subjective Quality of Speech Coders. *IEEE J. Sel. Areas Commun.* **1992**, *10*, 819–829. [CrossRef]
39. Armstrong, C.; Mckenzie, T.; Murphy, D.; Kearney, G. A Perceptual Spectral Difference Model for Binaural Signals. In Proceedings of the 145th Convention of the Audio Engineering Society, New York, NY, USA, 17–20 October 2018.
40. International Organization for Standardization. *ISO 226:2003, Normal Equal-Loudness-Level Contours*; ISO: Geneva, Switzerland, 2003.
41. Stevens, S.S. The Measurement of Loudness. *J. Acoust. Soc. Am.* **1955**, *27*, 815–829. [CrossRef]

42. Bauer, B.B.; Torick, E.L. Researches in Loudness Measurement. *IEEE Trans. Audio Electroacoust.* **1966**, *14*, 141–151. [CrossRef]
43. Bücklein, R. The Audibility of Frequency Response Irregularities. *J. Audio Eng. Soc.* **1981**, *29*, 126–131.
44. Baumgartner, R.; Majdak, P.; Laback, B. Modeling Sound-Source Localization in Sagittal Planes for Human Listeners. *J. Acoust. Soc. Am.* **2014**, *136*, 791–802. [CrossRef] [PubMed]
45. Algazi, V.R.; Avendano, C.; Duda, R.O. Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies. *J. Acoust. Soc. Am.* **2001**, *109*, 1110–1122. [CrossRef] [PubMed]
46. International Organization for Standardization. *ISO 389, Acoustics—Reference Zero for the Calibration of Audiometric Equipment*; ISO: Geneva, Switzerland, 2016.
47. Farina, A. Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. In Proceedings of the 108th Convention of the Audio Engineering Society, Paris, France, 19–22 February 2000.
48. American National Standards Institute. *ANSI S1.1-1994, American National Standard Acoustical Terminology*; American National Standards Institute: Washington, DC, USA, 2004.
49. International Telecommunications Union. *ITU-R-BS.1534-3, Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*; International Telecommunications Union: Geneva, Switzerland, 2015.
50. Mcgill, R.; Tukey, J.W.; Larsen, W.A. Variations of Box Plots. *Am. Stat.* **1978**, *32*, 12–16.